Electronic Letters on Computer Vision and Image Analysis 15(2):40-42, 2016

Hierarchical Visual Content Modelling and Query based on Trees

Arief Setyanto *+
* *School of Computer Science and ELectronics Engineering, University of Essex, Colchester, United Kingdom*
+ *School of Computer and Information Engineering AMIKOM, Yogyakarta,Indonesia*
Supervisors : John C Woods, Mohammed Ghanbari *

Received 1 July 2016; accepted 04 August 2016

Abstract

In recent years, such vast archives of video information have become available that human annotation of content is no longer feasible; automation of video content analysis is therefore highly desirable. The recognition of semantic content in images is a problem that relies on prior knowledge and learnt information and that, to date, has only been partially solved. Salient analysis, on the other hand, is statistically based and highlights regions that are distinct from their surroundings, while also being scalable and repeatable. The arrangement of salient information into hierarchical tree structures in the spatial and temporal domains forms an important step to bridge the semantic salient gap. Salient regions are identified using region analysis, rank ordered and documented in a tree for further analysis. A structure of this kind contains all the information in the original video and forms an intermediary between video processing and video understanding, transforming video analysis to a syntactic database analysis problem.

The contribution of this thesis [1] is the formulation of spatio-temporal salient trees the syntax to index them, and provides an interface for higher level cognition in machine vision.

In order to achieve those objectives, some works are carried out and presented in figure 1. Initial oversegmented supervoxels are obtained from watershed [2] and SLIC [3] algorithms. A merging procedure follows to [4] is performed to get a binary partition tree with a similarity measure considering supervoxels direction, mean colours and size. Merging order is determined by similarity measure that allows the most homogeneous supervoxels pair given highest priority. In every iteration, a pair of child nodes are merged and a new parent node is issued. Merging task is progressing until no more supervoxels pair available and a root node is achieved. A binary tree where all nodes represent a supervoxels in a number of consecutive frames is generated at this step.

Due to the numerous initial partitions, the tree is complex and contains a large number of nodes while few of them are important and correlated to the semantic objects. An evolutionary analysis, which is modified from [5] and [6], is carried out to detect a critical merging. The branch of the tree can be pruned in the critical nodes to get simpler version of BPT. An evaluation of three-level simplification is presented in [7]. The critical merging is indicated by the big gap between particular nodes with its merging result ($parent$ nodes), and it suggests that

Correspondence to: <arief_s@amikom.ac.id>

Recommended for acceptance by David Vázquez
DOI: http://dx.doi.org/10.5565/rev/elcvia.952
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

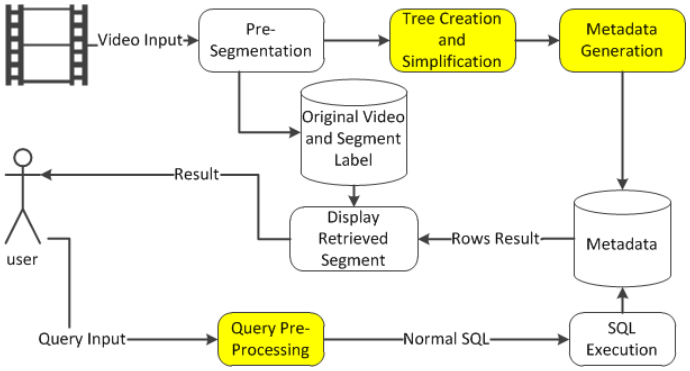

Figure 1: Building block of the work (Yellow blocks indicates the contribution of this thesis)

a pair of nodes may belong to different objects. Rank-ordered node in regard to the child-parent distances is composed, and the top rank salient nodes are evaluated against the ground truth objects.

The nodes in the tree contain some data aggregations such as mean colour, size, centroid position, neighbouring nodes and supervoxel direction. Video content metadata is generated as a documentation of the nodes of the binary partition tree. Some numerical attributes such as RGB code, motion directions are translated into human meaningful forms. Colour code is translated into a textual colour name, while the motion direction translated to direction name.

Metadata is designed to serve a visual information request through an SQL-like format. The metadata is stored and managed by the standard database management system. In order to enable spatio temporal visual information request, special keywords are introduced to deal with colour, motions and saliency-related query. Particular functions are dedicated to translate the special keywords to standard database syntax. The metadata provides an intermediate semantic level of content information. This enables a semantic request such as 'search football video with the red T-shirt' to be formulated in lower level form, for example 'search red moving object'. This work also enables to retrieve the most salient nodes in the BPT using 'top salient' keyword.

An evaluation is carried out based on the publicly available ground truth from xiph.org used by [8]. It shows that the object candidates in simplification result are close to the ground truth. The query retrieval for the top 10 salient nodes in the tree is also close to the ground truth objects. Some colour-related and motion-related query are demonstrated. It has shown that colour-related query obtains the best result from the initial supervoxels segmentation while the motion-related query gets a good result in the simplified tree.

References

[1] A. Setyanto, "Hierarchical Visual Content Modelling and Query based on Trees," Ph.D. dissertation, University of Essex, 2016. [Online]. Available: http://repository.essex.ac.uk/16903/

[2] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE transactions on pattern analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.

[4] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 9, no. 4, pp. 561–76, jan 2000.

[5] H. Lu, J. C. Woods, and M. Ghanbari, "Binary partition tree analysis based on region evolution and its application to tree simplification," *IEEE Transactions on Image Processing*, vol. 16, pp. 1131–1138, 2007.

[6] A. Setyanto, J. C. Wood, and M. Ghanbari, "Platform for Temporal Analysis of Binary Partition Tree," in *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013*, Poznan, Poland, 2013, pp. 45 – 50.

[7] A. Setyanto, J. C. Wood, and M. Ghanbary, "Evolution Analysis of Binary Partition Tree for Hierarchical Video Simplified Segmentation," in *Computer Science and Electronic Engineering Conference*, 2014, pp. 52–57.

[8] a.Y.C. Chen and J. Corso, "Propagating multi-class pixel labels throughout video frames," in *Image Processing Workshop (WNYIPW), 2010 Western New York*, 2010, pp. 0–3.